

**PATENT**  
**5760-12900**  
**VRTS 0421**

"EXPRESS MAIL" MAILING LABEL NUMBER  
EV318247881US

DATE OF DEPOSIT: SEPTEMBER 30, 2003

I HEREBY CERTIFY THAT THIS PAPER OR  
FEE IS BEING DEPOSITED WITH THE  
UNITED STATES POSTAL SERVICE  
"EXPRESS MAIL POST OFFICE TO  
ADDRESSEE" SERVICE UNDER 37 C.F.R.  
§1.10 ON THE DATE INDICATED ABOVE  
AND IS ADDRESSED TO THE  
COMMISSIONER FOR PATENTS, P.O. BOX  
1450, ALEXANDRIA, VA 22313-1450

  
Derrick Brown

**COMPUTER SYSTEM AND METHOD FOR  
PERFORMING LOW IMPACT BACKUP OPERATIONS**

By:

Robert P. Rossi

Atty. Dkt. No.: 5760-12900

B. Noël Kivlin /CLJ  
Meyertons, Hood, Kivlin, Kowert & Goetzel, P.C.  
P.O. Box 398  
Austin, TX 78767-0398  
Ph: (512) 853-8800

## BACKGROUND OF THE INVENTION

### Field of the Invention

[0001] This invention relates to the field of computer processing and, more particularly, to the performance of data management processes such as backup operations on computer systems.

### Description of the Related Art

[0002] As computers become more integral to modern business and industry, maintaining constant availability and maximizing responsiveness assumes a greater importance. A trend is developing whereby computer systems that have traditionally been used primarily during only a portion of each day, such as business hours, may now be in high demand throughout the day and night, everyday (so-called 24x7). Furthermore, it may be difficult to predict in advance when system demand may be high or low.

[0003] Various data management tasks, such as data backup, virus detection, and search indexing, have traditionally been scheduled late at night or on weekends to minimize interference with user applications. However, due to the increased demand for computer system resources during previously “off-peak” hours, it has become more difficult to reliably schedule these data management tasks when demand is low.

[0004] Furthermore, since many data management tasks are designed to complete their desired operations very quickly, they may consume large amounts of system resources during the time they are running. The performance of other applications may thus be negatively impacted if they are concurrently run with the data management tasks. For example, many data backup applications are designed to minimize the length of time a backup operation requires by transferring data to a backup medium at very high transfer

rates. However, the I/O bandwidth consumed by such operations may noticeably interfere with other applications that also require I/O access at the same time.

[0005] As illustrated in Fig. 1, for example, a backup session associated with a backup application may be scheduled to execute once a day, at the same time every day. Resource utilization due to the scheduled backup sessions is illustrated by the solid line. As depicted by the dashed line, resource utilization by one or more user applications may fluctuate gradually over time. Unfortunately, as the figure also illustrates, the backup session may be scheduled at times when demand for the I/O resources by the user applications is also high. This may therefore result in decreased performance from the user's perspective.

## SUMMARY OF THE INVENTION

5 [0006] Various embodiments of a system and method for low impact backup are disclosed. In one embodiment, a method may comprise monitoring utilization of a system resource and a data management process selectively performing I/O operations dependent upon the monitored utilization of the system resource. The data management process may include functionality to backup desired data from a storage medium to a backup medium. In one particular implementation, the I/O operations may be allowed to be performed in  
10 response to the utilization of the system resource falling below a predetermined threshold.

[0007] In another embodiment a method may comprise performing a plurality of I/O operations to complete a data management process executed by an application. The application separates said plurality of I/O operations with intermittent delays to achieve  
15 time-slicing of the data management process with respect to one or more other applications.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Fig. 1 is a graph illustrating I/O resource demands due to user application workload and backup session workload.

5

[0009] Fig. 2 is a block diagram of one embodiment of a computer system.

[0010] Fig. 3 is a flowchart illustrating one embodiment of a method for low impact backup.

10

[0011] Fig. 4 is a graph illustrating exemplary I/O resource demand due to user application workload and backup session workload as a result of an implementation of the low impact backup method of Fig. 3.

15 [0012] Fig. 5 is a flowchart illustrating another embodiment of a method for low impact backup.

[0013] Fig. 6 is a graph illustrating exemplary I/O resource demand due to user application workload and backup session workload as a result of an implementation of the low impact backup method of Fig. 5.

20

[0014] Fig. 7 is a graph illustrating exemplary I/O resource demand due to user application workload and data management task workload as a result of another implementation of a low impact backup method.

25

[0015] While the invention is susceptible to various modifications and alternative forms, specific embodiments are shown by way of example in the drawings and are herein described in detail. It should be understood, however, that drawings and detailed description thereto are not intended to limit the invention to the particular form disclosed,

but on the contrary, the invention is to cover all modifications, equivalents and alternatives falling within the spirit and scope of the present invention as defined by the appended claims.

## DETAILED DESCRIPTION

[0011] Turning now to Fig. 2, a block diagram of one embodiment of a computer system 200 is shown. Computer system 200 includes a processor 204 coupled to a main memory 202. Processor 204 and main memory 202 are in turn connected to an I/O subsystem 210, which comprises an I/O interface 212, a hard disk drive 214, a network interface 216, and a removable storage 218. It is noted that computer system 200 may be representative of a laptop, desktop, server, workstation, terminal, personal digital assistant (PDA) or any other type of computer system.

10

[0012] Processor 204 is representative of any of various types of processors such as an x86 processor, a PowerPC processor or a SPARC processor. Similarly, main memory 202 is representative of any of various types of memory, including DRAM, SRAM, EDO RAM, Rambus RAM, etc.

15

[0013] I/O interface 212 is operational to transfer data between processor 204 and/or main memory 202 and one or more internal or external components such as hard disk drive 214, network interface 216 and removable storage 218, as desired. For example, I/O interface 212 may embody a PCI bridge operable to transfer data from processor 204 and/or main memory 202 to one or more PCI devices. I/O interface 212 may additionally or alternatively provide an interface to devices of other types, such as SCSI devices and/or Fibre channel devices.

[0014] Hard disk drive 214 may be a non-volatile memory such as a magnetic media. Network interface 216 may be any type of network adapter, such as Ethernet, fiber optic, or coaxial adapters. Removable storage 218 is representative of a disk drive, optical media drive, tape drive, or other type of storage media, as desired.

25

5 [0015] In addition to the various depicted hardware components, computer system 200 may additionally include various software components. For example, Fig. 2 illustrates an operating system 250 and one or more user applications 260 stored in main memory 202. Operating system 250 is representative of any of a variety of specific operating systems, such as, for example, Microsoft Windows, Linux, or Sun Solaris. As such, operating system 250 may be operable to provide various services to the end user and provide a software framework operable to support the one or more user applications 260. The one or more user applications 260 are representative of any software programs which may be executable on computer system 200.

10

[0016] A backup application 270 is also shown stored in main memory 202. Backup application 270 may be operable to create a backup copy of selected data from storage such as hard disk drive 214 to a backup medium such as removable storage 218. Alternatively, selected data may be transferred from hard disk drive 214 to network interface 216 for transmission to and backup in a remote system.

15 [0017] As will be described in further detail below, backup application 270 may be configured to implement various low impact backup methods. In one embodiment, backup application 270 may be operable to control the execution of a backup session dependent on the utilization of I/O bandwidth or another monitored system resource(s). In another embodiment, backup application 270 may be operable to time slice a backup session into separate execution periods to minimize the impact on one or more user applications 260 which may be running concurrently.

25 [0018] Fig. 3 is a flowchart illustrating one embodiment of a low impact backup method. Referring collectively to Figs. 2 and 3, in 300, a user may set a series of parameters that control various aspects of the method, as described below. In 302, backup application 270 monitors the utilization of I/O subsystem 210 (directly or indirectly) which may be a result of execution of the one or more user applications 260 (or due to



other consumers of I/O subsystem 210). More particularly, in one embodiment backup application 270 may determine the amount of bandwidth being consumed in relation to the interfaces between I/O interface 212 and hard disk drive 214, network interface 216 and/or removable storage 218. It is noted that in other embodiments, other aspects  
5 associated with the utilization of I/O subsystem 210 may be additionally or alternatively monitored, such as the number, type, and/or size of I/O requests being provided to I/O interface 212 during a selected period of time.

[0019] The specific manner in which resource utilization is monitored in step 302  
10 may vary depending upon the particular embodiment and the specific resource being monitored. For example, in one embodiment backup application 270 may access a utility provided by operating system 250 that functions to provide an indicator of utilization of a monitored resource, such as I/O subsystem 210. Alternatively, backup application 270 may be operable to directly query hardware in computer system 200, or to query drivers  
15 associated with such hardware. In yet another embodiment, backup application 270 may be operable to interact with a hardware monitor that monitors usage of a system resource.

[0020] In 304, backup application 270 determines whether the I/O utilization as determined in step 302 satisfies a trigger threshold. In one particular implementation, the  
20 trigger threshold may be specified by one or more of the parameters set in step 300. For example, the trigger threshold may be set by a user such that it is satisfied when the utilization of the total possible bandwidth between hard disk drive 214 and removable storage 218 falls below a certain percentage.

25 [0021] If in step 304 it is determined that the trigger threshold has been satisfied, backup application 270 executes a backup session in step 306. The backup session may include backing up of the entire contents or selected contents of hard disk drive 214 (or other storage device), as desired. It is noted that by executing a backup session when the I/O utilization (e.g. bandwidth utilization of hard disk drive 214 and/or removable media

218) is relatively low, the performance impact on user applications 260, which may be running concurrently, may be negligible.

5       **[0022]**     If in step 304 it is instead determined that the trigger threshold has not been satisfied, backup application 270 may determine in step 308 the amount of time which has elapsed since the last backup session was performed. If the time since the last backup session is less than a predetermined interval, backup application 270 may return to step 302. It is noted that in this embodiment the trigger threshold will not be satisfied when utilization of the I/O subsystem 210 is relatively high, which may be the result of  
10     operations invoked by user applications 260. Since backup operations in step 306 are not performed in this situation, further loading of system resources such as I/O subsystem 210 may be avoided, and application performance from a user's perspective may be improved.

15       **[0023]**     If the time since the last backup session in step 308 is greater than the predetermined interval, backup application 270 may instead advance to step 306, in which a backup session is executed. In this manner, the time between successive backup sessions may be guaranteed to not exceed the predetermined interval, regardless of whether I/O conditions are favorable or not. Such functionality may be desirable to ensure that data is backed up with at least some frequency. It is noted that in one particular  
20     implementation, the predetermined interval may be specified by one or more of the parameters set in step 300. It is also noted that in alternate embodiments, the operations of step 308 may be omitted in such a way that a backup session may be invoked only if the trigger threshold in step 304 is satisfied (i.e., if resource utilization is relatively low).

25       **[0024]**     In a further implementation of the low impact backup method illustrated by Fig. 3, a set amount of time may be required to elapse (following initiation or completion of a given backup session) before another backup session can be initiated in step 306. This may thus prevent consecutive backup sessions from being triggered too frequently.

[0025] In addition, in a further implementation of the low impact backup method illustrated by Fig. 3, the I/O utilization (or other system resource utilization) as determined in step 302 may be required to satisfy a given threshold for at least a given period of time before a backup session will be invoked in step 306. Inclusion of such  
5 functionality may thus avoid initiating a backup session in response to short sporadic breaks in system demand.

[0026] It is noted that, in various alternative embodiments, backup application 270 may monitor other system resources, such as the usage of processor 204, to determine if a  
10 trigger threshold has been satisfied. In addition, any various combinations of monitored resources may be used to determine if a trigger threshold has been satisfied.

[0027] Still further, in yet an additional alternate embodiment, various aspects associated with the utilization of I/O subsystem 210 (or other system resources) may be  
15 monitored and weighted differently (or prioritized) in step 304 to determine whether a trigger threshold has been satisfied. For example, write activity to hard disk drive 214 may be much more important than read activity in limiting a backup session. Thus, the bandwidth consumed by write activity may be given a higher weighting when calculating if the trigger threshold has been satisfied.

20

[0028] It is also noted that, in various embodiments, the parameters set in step 300 may further control, for example, which system resource or combination of resources may be monitored in step 302, how those resources will be weighted in relation to each other, and how long system resources must remain below threshold levels for the trigger  
25 threshold to be satisfied.

[0029] Fig. 4 is a graph illustrating an exemplary result which may be attained through implementation of the low impact backup method of Fig. 3. Comparing Fig. 4 to Fig. 1, the same demand for I/O bandwidth by user applications 260 (as shown by the

dashed line) may be presented to the system. However, in Fig. 4 the low impact backup method of Fig. 3 triggers backup sessions such that they are performed when user demand is relatively low, thereby minimizing the interference between backup operations and user application workload.

5

**[0030]** While in the embodiments described above the functionality of Fig. 3 is implemented by a backup application, other embodiments are possible. For example, in one alternative embodiment, the low impact backup method of Fig. 3 may be carried out by functionality within operating system 250 as part of a backup utility.

10

**[0031]** Fig. 5 is a flowchart illustrating an alternative embodiment of a low impact backup method. As will be described below, Fig. 5 illustrates a method in which a backup session is carried out by time slicing, i.e. using backup application 270 to control the execution of a backup session such that data is backed up intermittently over a period of time. By intermittently backing up data associated with a backup session, interference with user applications which may be running concurrently may be less noticeable to a user.

**[0032]** As illustrated in Fig. 5, in step 500, a user may set a series of parameters that control various aspects of the method, as will be described below. In 502, backup application 270 executes a first backup session slice. A backup session slice is a portion of a backup session which is separated from other backup session slices by an intermittent time delay. In one embodiment, a backup session slice may be specified by the amount of data transferred between time delays. For example, 512 kilobytes of data may be transferred to a backup medium every 500 milliseconds. In another embodiment, a backup session slice may be specified by a length in time during which I/O operations for backup are performed. For example, a computer may perform backup activity for 500 milliseconds, and then wait another 1500 milliseconds to perform another backup session slice. These parameters may be set in step 500, as described above.

[0033] In 504 backup application 270 determines if the backup session is complete, i.e. if the last slice of the backup session has been backed up to a backup medium such as removable storage 218. If the backup session is complete, the method ends. Alternatively, 5 if the backup session is incomplete, backup application 27 advances to step 506, wherein backup application 270 delays execution of I/O operations associated with the next backup session slice for an intermittent delay value. In one embodiment, the intermittent delay value may be set by the user in step 500. Backup application 270 may then return to step 502.

10

[0034] Fig. 6 is a graph illustrating an exemplary result which may be attained through implementation of the low impact backup method of Fig. 5. Comparing Fig. 6 to Fig. 1, the same demand for I/O bandwidth by user applications 260 (as shown by the dashed line) may be presented to the system. However, in Fig. 6 the low impact backup 15 method of Fig. 5 time slices I/O operations associated with each backup session such that they are staggered over a period of time. It is noted that by time slicing a backup session, the impact of the backup operations on user applications 260 which may be running concurrently may be reduced.

20 [0035] In another embodiment of the embodiment described in Fig. 5, the intermittent delay provided in step 506 between backup session slices may be varied depending upon a utilization of a monitored system resource such as I/O subsystem 210, as described in step 302 of Fig. 3. In such an embodiment, backup session slices may be clustered closer together when demand for system resources is low, and spread further apart as demand 25 for system resources is high.

[0036] It is noted that various alternate embodiments may combine the operations of Figs. 3 and 5 in a variety of ways. For example, the resource demand monitoring methodology of Fig. 3 might be combined with the time-slicing methodology of Fig. 5. In

one such embodiment, each backup session is triggered according to demand for system resources, such as I/O subsystem 210. Once triggered, the backup session would then be time-sliced, in effect executing steps 502, 504 and 506 of Fig. 5 during step 306 of Fig. 3. This embodiment may thereby further minimize interference between the backup session and user applications 260. It is noted that such an embodiment may result in triggering a backup session when the demand for system resources is low, and in further distributing the impact of a backup session across a period of time.

[0037] In a further implementation of the embodiment described above, the delay provided between execution of backup session slices may be varied according to a utilization of a monitored system resource such as I/O subsystem 210. Fig. 7 is a graph illustrating an exemplary result which may be attained through implementing such an embodiment. Accordingly, backup session slices may be both triggered and clustered together when demand for system resources is low, and spread apart when demand is high, thereby providing an optimal utilization of otherwise unused resources and minimizing interference with user applications.

[0038] It is further contemplated that, in various alternative embodiments of the low impact backup methods described above, historical information may be used to control the performance of the backup sessions and/or backup session slices. The historical information may include, for example, information indicative of a prior history of resource utilization within the computer system and/or a prior history of the performance of backup sessions and/or backup session slices. Based upon the historical information, one or more of the parameters as discussed in the above embodiments (e.g., the trigger threshold level, how long system resources must remain below threshold levels for a trigger to be satisfied, which system resources are used determine whether a trigger threshold has been satisfied, and/or the clustering of backup session slices) may be automatically tuned. In alternative embodiments, future backup sessions may be controlled based upon the historical information.

[0039] For example, in one embodiment, the low impact backup method may be configured to determine from the historical information that backup sessions and/or backup session slices have not been invoked as often as desired, or that a trigger threshold  
5 has not been satisfied as frequently as desired. Accordingly, in response to such a determination, the low impact backup method may automatically lower the trigger threshold. In another embodiment, the low impact backup method may be configured to determine from the historical information that utilization of a monitored system resource is typically relatively low at a particular time of a day (or of multiple days). In response  
10 to such a determination, the low impact backup method may automatically invoke a backup session at the particular time on a subsequent day.

[0040] It is noted that in other embodiments, rather than performing I/O operations associated with a backup application in steps 306 and/or 502, I/O operations to perform  
15 other types of data management tasks may be performed. For example, I/O operations associated with data management tasks such as virus scanning or search indexing may be triggered according to monitored system resources or time sliced to minimize the impact of such data management tasks on user applications 260.

20 [0041] It is noted that any of the embodiments described above may further include receiving, sending or storing instructions and/or data that implement the operations described above in conjunction with Figs. 3 – 7 upon a computer readable medium. Generally speaking, a computer readable medium may include storage media or memory  
media such as magnetic or optical media, e.g. disk or CD-ROM, volatile or non-volatile  
25 media such as RAM (e.g. SDRAM, DDR SDRAM, RDRAM, SRAM, etc.), ROM, etc. as well as transmission media or signals such as electrical, electromagnetic, or digital signals conveyed via a communication medium such as network and/or a wireless link.

**[0042]** Although the embodiments above have been described in considerable detail, numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.